UNFAIR INEQUALITY MEASUREMENT IN CHINA AND SOUTH AFRICA

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Xinyue Pei

May 2019

ABSTRACT

This thesis applied two ways to measure unfair inequality in China and South Africa. Firstly, we focus on the unfair inequality violating the principle of Equal Opportunity (EOp), which answers to what extent individuals income difference are due to factors beyond their control (in Roemer's terminology "circumstances"). Utilizing the China Family Panel Studies Survey, we measure this Roemerian unfair inequality in individual earnings in aggregate and for each of 10 birth cohorts from 1955-1985. The aggregate result shows that Roemerian unfair inequality takes up nearly 24% of the overall inequality. The cohort pattern shows an increasing trend of Roemerian unfair inequality for the younger cohorts. Among all circumstance variables, gender is the most influential one, contributing to nearly half of the unfair inequality. But the impact of gender decreases for the later cohorts. In South Africa, we use the National Income Dynamics Study (NIDS) to measure this Roemerian unfair inequality in individual annual gross income. On average, Roemerian unfair inequality takes up nearly 16% of the overall inequality. This unfair inequality ratio shows an increasing trend for the younger cohorts. Among all circumstance variables, parental education and race are the most influential ones, both contributing to nearly 23% of the unequal opportunity.

Second, we reconcile both the principle of Equal Opportunity (EOp) and Freedom from Poverty (FfP) to measure unfair inequality (Hufe, Kanbur, & Peichl, 2018). In China, HKP unfair inequality takes up 27% of the outcome inequality and remains relatively stable over the cohorts. Further decomposing HKP unfair inequality, we find that the stability results from a combination of

an increasing trend of inequality violating EOp and decreasing trend of inequality violating FfP. In South Africa, the HKP unfair inequality takes up 21% of the outcome inequality. After the decomposition, we find an increasing share of inequality violating FfP for the younger cohort. Finally, we tried to include 2560 individuals with non-positive income in CFPS dataset by using a concave log-like transformation and repeat our two measurements above. With this adjustment, we found significant upward corrections of unfair inequality.

ACKNOWLEDGEMENTS

I would first like to thank my advisor Professor Ravi Kanbur for your patience, motivation and immense knowledge. I could not thank you enough for guiding my studies in inequality and poverty topics that are greatly eye-opening and in-depth. The chance of visiting South Africa given by you is something that I will remember for the life time. You have had an incredible influence on not only my academic career but also my life. Thank you Professor Nancy Chau for your brilliant comments and suggestions whenever I ran into a question about my research. As my committee member and mentor, you have taught me more than I could ever give you credit for here. I would also wish to express my gratitude to Professor Kaushik Basu for your professional and personal support. Working with you was such an extraordinary and inspiring experience. Thank you professor Chris Barrett and Professor Douglas Miller for reading through my paper and giving me valuable advice. Thank you Bridget Cristelli for being the best program assistant.

To my friends: Mingjie Zhu, Zekun Ma, Yuyi Chen, Elaine Qiu, Yujie Feng, Zihan Hu, Youwei Yang, Amber Liu, Edith Yuan, Karina Acosta, Ali Abbas, Prankur Gupta, Sasha Tettenhorst, thank you for always having an open door and mind. You have given me all the reasons to carry on.

Most importantly, none of this could have happened without my family. Thank you, my mom and dad, for offering encouragement through phone calls and texts every week. Your unconditional support keeps me sane and courageous.

Finally, Tyler Porter, thank you for always being there and having my back through the good and the bad.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Since the reform and opening up institutional change in the late 1970s, China's record in poverty reduction and economic growth is perhaps without parallel in human history. In less than four decades, China lifted 850 million people out of poverty, with an economy expanded at an annual average 9.5% growth rate and increased almost 35 times in size . Alongside the spectacular achievement, equally remarkable is the sharp rise in income inequality[1]. Economic growth in South Africa has also a striking uneven feature. In 2014, the Gini coefficient reached 0.65 in expenditure data, and 0.69 based in income. [2]

However, calls for reducing inequality has never received unconditional support. Outcome inequalities are recognized as necessary to individual incentives and to achieve justice in a market economy. Beginning with Rawls (1958, 1971), political philosophers and economists have made effort to distinguishing differentiate between fair = and unfair inequalities by bringing "individual responsibility". They pointed out that inequalities are unfair if they are originated from factors beyond individual control. A prominent field of research in this area is Equality of Opportunity (EOp) which builds on the early ideas of Cohen(1989) Fleurbaey (1995), and Roemer (1993). This literature defines determinants of individual outcomes as two categories – circumstances and efforts. Circumstances lies beyond the control of an individual (for example, sex, race, birth place, socioeconomic status of parents). To the contrary, efforts can be held within individual responsibility and should be rewarded. Society is considered less fair if circumstances explain a large portions of the individual' actual out-

[1] According to data source from Penn World Table (Version 7.1), the Gini coefficient rockets from 0.26 to 0.43 between 1980 and 2010.

[2] Data from Statistics South Africa

come.

Nonetheless, the EOp principle itself is insufficient to define fairness. Another important principle – Need Principle – calls for the equal satisfaction of basic needs as an essential part to define fairness. Bourguignon and Ferreira (2007) also define Equity using two these principles: 1) equal opportunity and 2) avoidance of extreme deprivation in outcomes. With this broader definition for fairness, Hufe, Piechl and Kanbur (2018) enriched the empirical implementation of unfair inequality by incorporating the principles of Equality of Opportunities (EOp) and Freedom from Poverty (FfP). Their results show that unfair inequality is greater in the US than anywhere in Europe, and that it has been increasing over time. The findings also show that relying solely on EOp principle will largely underestimate unfair inequality.

Despite the fact that there now is a burgeoning literature that incorporates the idea of fairness into inequality research, the number of empirical applications that explicitly apply this concept under the context of China or South Africa is scarce.

Therefore, the main focus of this paper is to provide a more comprehensive analysis of unfair income inequality in China and South Africa by distinguishing two levels of unfair inequality we mentioned above: The first one follows Roemers Equal Opportunity (EOp) tradition, where we estimate parametric models using robust OLS regression in order to quantify the role of ethnicity, gender, parents' education and occupation, region of origin in generating inequality in current income; the second method focuses on measuring the divergence between the real-world income distribution with a constructed normative income distribution that follows both the principle of EOp and Freedom from Poverty (FfP). We utilize CFPS a national representative dataset to carefully quantify some features of these two levels of unfair inequality for people who were born between 1955 and 1985 . Our results indicate unfair inequality takes up around of the overall inequality (24% for the first method and 27% for the second method). Gender dominates all other variables by explaining more than half of the unequal opportunity. The younger cohorts experience more unfair inequality violating EOp and less unfair inequality violating FfP. At last, we re-estimate the unfair inequality by using the same methodologies but including the non-positive incomes a commonly omitted part in most of the empirical implementations. We find that unfair inequality increase 2 times in our first measure and 8 times in our second measure, which indicates a huge under-estimation effect. The structure of the paper is the following: Chapter 1 gives an philosophical and economic overview of the unfair inequaluty and its measurement, and elaborates the methods of measuring unfair inequality in the framework of the Roemer model as well as the Hufe, Piechl and Kanbur model. Chapter 2 is an empirical application in China. We present CFPS data and measure unfair inequality using the two models described in the previous section. Due to a large number of 0 incomes in the CFPS dataset, we also apply a concave log-like transformation (Ravallion, 2017) to accommodate non-positive income and re-measure the unfair inequality. Chapter 3 applies the same measurement methods in the context of South Africa using NIDS dataset.

CHAPTER 1

UNFAIR INEQUALITY MEASUREMENT

1.1 Literature Review

The Unfair and Unfair inequality framework originates from a debate in political philosophy beginning with John Rawls (Rawls, 1971). This debate inspired a new approach to egalitarianism by questioning the equality of welfare of utility of being unrealistic and unethical, as it fails to recognize the role of individual responsibility (Dworkin, 1981; Fleurbaey, 1995; J. Roemer, 1998). In the last two decades, the inequality of opportunity literature has flourished after influential contributions by Roemer (J. Roemer, 1998) and Fleurbaey (Fleurbaey, 2008). This literature defines determinants of individual achievements as two categories - *'circumstances'* (genetic endowments, socioeconomic background, etc.) and *'individual effort'*. Furthermore, inequalities due to circumstances are unfair and should be eliminated, while inequalities due to unequal effort should be acceptable. (Bourguignon, Ferreira, & Menéndez, 2007; Lefranc, Pistolesi, & Trannoy, 2008). By partitioning the population into 'types' with identical circumstances, we can and measuring the divergence of the mean outcome levels across types (Van De Gaer, 1995). In other words, this measure of unfair inequality is similar to between-type inequality but only based on the mean levels for each type.

Bases on the this criterion, a broad range of empirical implementations of unfair inequality measurements have burgeoned. Among them, two categories are most commonly used. One is the parametric method and the other is the non-parametric method. Bourguignon et al. (2007) formalized the parametric method first, which soon got advanced and improved by Ferreira and Gignoux (2008). This method requires some functional form assumptions about outcome on circumstance and effort. A smoothed distribution conditional on circumstances is generated and used for measures of unfair inequality. Checchi and Peragine (2010a) developed two non-parametric methods, which did not require functional form assumptions. The first one is called the ex-ante approach, which partitioned the population into groups of individuals with the same circumstances and call these groups 'types'. The total inequality in earnings is then decomposed into between-type (unfair) and within-type (fair) components. Unfair inequality is estimated as between-types earning differentials. The second one is called ex-post approach, which partitioned the population into groups of people that yield same level of effort and call these groups 'tranches'. Similarly, total inequality now can be decomposed into the within-tranche and between-tranche part. Unfair inequality is estimated as within-tranches earning differentials. The ex-ante and ex-post approaches offer different estimate results by capturing different aspects of unfair inequality. For example, they reflect unfair inequality in the entry of pre-labor market and post-labor market correspondingly. (Checchi, Peragine, & Serlenga, 2010b)

The first measurement method this paper uses is based on the ex-ante approach due to the following two reasons. Firstly, our interest lies in how much of the inequality can be explained due to circumstances. Secondly, the ex-post approach needs to identify individuals with same degree of effort. This requires a rich data set that allows the the within group population to be divided into percentiles. It is also hard to define variables that are purely considered as "effort" when the dataset is not affluent enough. Therefore, it is more appropriate for us to implement the ex-ante approach given the absence of big datasets available in China or South Africa, since their data contained more limited information comparing to developed countries like US or OECD countries.

Most recently, researchers start to argue that this weak criterion is not the only moral intuition behind peoples concern with unfair income distribution. Hufe, Peichl and Kanbur (2018) admit that income variation associated with circumstance is illegitimate, which they call disobeying criterion Equality of Opportunity (EOp). Moreover, fair income distribution should also satisfy the criterion of Freedom from Poverty (FfP), which requires peoples income to reach the poverty line at least, regardless of their circumstance or effort. Based on the ex-ante non-parametric method, they first partition the population according to the circumstances and then build a fair income distribution that satisfies both of the criterions above. This fair distribution requires moving all those who are 'poor' (below the poverty line) up to the poverty line to satisfy FfP. Then, to achieves the other principle EOp they focus on the people who are 'rich' (above the poverty line). They partition people 'types' according to their circumstance and equalize mean incomes across the 'types' by proportionate scaling up or down without changing the overall mean income. After the ideal norm distribution is constructed, the distance metric between the ideal and the observed distribution measures unfair inequality. They applied this measurement in 31 European countries and a longitudinal data for the US, in which they found sizable upward corrections of unfair inequality.

This paper also applies this HKP unfair inequality measurement in China and South Africa. This will be the first attempt both in China and South Africa context. The next two subsections provides detailed explanation of the two methodology applied in this study– Roemer model and HKP model.

1.2 Roemerian Unfair inequality

The Roemerian unfair inequality is base on parametric method. (Bourguignon, Ferreira, & Menéndez, 2007; Ferreira & Gignoux, 2008). We consider a stylized model of outcome (in our context, annual income) of the form $y_{it} = f(C_i, E_{it}, \epsilon_{it})$,

$$y_{it} = f(C_i, E_{it}, \epsilon_{it}) \tag{1.1}$$

where the variable of interest y_{it} depends on exogenous circumstances C_i as well as effort E_{it}. We further assume circumstances are time-invariant variables and efforts are time-variant variables. Moreover, E_{it} is endogenous and can be determined by C_i , which implies that the model can be expressed as

$$y_{it} = f(C_i, E(C_i, v_{it}), \epsilon_{it}) \tag{1.2}$$

If we are mainly interested in the impact of circumstances as our unfair inequality, then estimating the channel directly through effort is not necessary. (Bourguignon, Ferreira, & Menéndez, 2007; Trannoy, Tubeuf, Jusot, & Devaux, 2010) Hence, the estimates δ in the reduced form (1.3) suffices to measure the overall direct and indirect impact of observable circumstances.

$$y_{it} = \delta C_i + \epsilon_{it} \tag{1.3}$$

We can then predict a counterfactual earnings $\tilde{y}$ that is only determined by circumstances. In the case of absolute EOp, there is no difference in income due to (observed) circumstances C_i. Consequently, unfair inequality can then be measured as the inequality of these counterfactual earnings levels, where differences are only due to differences in circumstances. For a given inequality measurement I , the counterfactual incomes are used to estimate the Absolute

Inequality of Opportunity $IOA = I(\tilde{y})$ The share of IOA in the overall inequality is called Relative Inequality of Opportunity $IOR = \frac{I(\tilde{y})}{I(y)}$.

1.3 HKP Unfair inequality

The measure of HKP unfair inequality follows the non-parametric fashion which relies on the norm-based approach. Below we will provide a succinct illustration. For more detail, we ask readers to refer to Hufe et.al (2018). We begin with a society with finite population $N = \{1, 2, \ldots, n\}$ with non-negative incomes $Y = \{y_1, y_2, \ldots y_n\}$. We further assume a set of circumstance $\Omega = \{C^1, C^2, \ldots C^n\}$ and effort θ which is a scalar variable. They jointly determine the outcome y. Of all circumstances $C^k \in \Omega$, we can partition N into a set of mutually exclusive types: $T = \{t_1, t_2, \ldots t_m\}$. The HKP method requires the construction of $Y^r = \{y_1^r, y_2^r, \ldots y_n^r\}$ that satisfies a set of normative requirement to meet both the principle of EOp and FfP. Denote D as the set of all possible income distributions, the restrictions on D are as follows:

(1) Constant resource

$$D^1 = \{D : \sum_i y_i^r = \sum_i y_i\} \tag{1.4}$$

which implies the total amount of resources should match its empirical counterpart.

(2) Equal Opportunity (EOp)

$$D^2 = \{D : \mu_t^r = \frac{1}{N_t} \sum_{i \in t} y_i^r = \frac{1}{N} \sum_i y_i = \mu, \ \forall t \in T\} \tag{1.5}$$

which requires equality across types with respect to their average outcomes.

(3) Freedon from Poverty (FfP)

$$D^3 = \{D : y_i^r = y_{min}, \forall i \in P\} \tag{1.6}$$

which claims that the poor iP all have an equal claim to nothing less than exactly the subsistence level.

(4) In order to avoid making the rich iR below the poverty line

$$D^4 = \{D : y_i^r > y_{min} \ \ \forall i \in R\} \tag{1.7}$$

(5) To make sure for the rich, the within-type possible distributions for excess income remains unaltered:

$$D^5 = \{D : \forall t \in T, \frac{y_i^r - y_{min}}{y_j^r - y_{min}} = \frac{y_i - y_{min}}{y_j - y_{min}} \ \forall i, j \in t \cap R\}. \tag{1.8}$$

The intersection of these subsets, $\cap_{s=1}^{s=5} D^S$ yields a singleton which defines Y^r. Details of the formula for Y^r is provided in the Appendix A. After the population is characterized by a vector of income pairs of observed income and ideal income $\{(y_1, y_1^r), (y_2, y_2^r), \ldots (y_n, y_n^r)\}$, unfair inequality is now measured by summarizing the differences between Y and Y^r. The next question is to measure the divergence between the two. To ensure close proximity and comparability to the literature on the measurement of inequality of opportunity (Ferreira & Gignoux, 2008; Hufe et al., 2018), we continue to use mean log deviation (MLD).

1.4 Discussion and future work

The Roemerian parametric method and HKP non-parametric method has some common benefits: Especially in developing countries with less national surveys providing detailed circumstance variables, these two methods make it possible and easy to estimate unfair inequality. All we need to do for the Roemerian parametric method is to estimate the reduced model using variables such as gender, race, region of birth or some parental status variables (income, occupation or education). And for the HKP method, we can use the same circumstances as above and partition the population into types. One additional information we need is the poverty line, which is available most the time.

In addition, they have their own distinctive advantage and can be a good complement to each other. In order to keep enough observations within a type, the HKP method requires large sample sizes or a small number of types, otherwise the estimates is not representative or reliable. Therefore, the circumstance variables selected in the HKP model is usually limited. Roemerian parametric method overcomes this problem by using a linear regression. Due to its parametric fashion, Roemer method can be used to estimate the contribution of each observed C_i variable. A disadvantage for the parametric method is obvious its aspect of fairness is limited to only the principle of EOp. Hence, the unique contribution of HKP method is that it combines both the principle of EOp and FfP in the measurement of unfair inequality by treating both as co-equal roles.

However, for both of the models, omitted variables are widespread. The observable circumstances set is clearly a subset of all the circumstances. Even if we have a perfect dataset that includes all the aspects of an individual, which variable to include as circumstance is also a controversial economic and philosophical question. For the parametric Roemer model, the coefficients will be biased and cannot be interpreted as causal due to the omitted variable. Moreover, the relation no longer stands for a nonlinear model. For the non-parametric HKP model, the omitted variable problem is reinforced. Since we want each type cell to include enough observations, we have to limit the circumstance variables we choose. Ferreira and Gignoux (2008) have argued that the reduced form will lead to a lower bound estimate of circumstances. Yet the interpretation of this lower bound can also be a problem in policy inference. (Kanbur & Wagstaff, 2014). In spite of their weaknesses, these two methods are good starting points for any study into unfair inequalities, especially in China where the related research is quite rare.

Other than the two method mentioned above, the empirical literature on unfair inequality keep flourishing. For example, Niehues and Peichl (2014) proposed an upper-bound estimate method of inequality of opportunity. This upper bound measurement requires rich panel data. If all circumstances are assumed as time-invariant, inequality associated with individual fixed effects can provide an upper-bound estimate. In other words, they let the individual idiosyncrasy term and to capture the inequality of opportunity. They found this upper bound measurement yields significant higher estimates compare to the traditional Roemer Model. Even if some effort variables are also time-invariant, then that estimate is upwardly-biased hence still an upper bound. However, this method overlooks the possibility of having time-varying circumstances, for example, later change of family composition and luck. Some other literature focus more on the assumptions about the functional form specifications $y_{it} = f(C_i, E_{it}, \epsilon_{it})$. Instead of simply throwing all the circumstances variables in the reduced form, Brunori et al. (?) tried to select the most suitable specifications using the technique of machine learning such as conditional inference random forests (2006).

Although the literate of unfair inequality is gaining more attention, it has not been fully incorporated into the mainstream economics studies yet. Ferreira and Peragine (Ferreira & Peragine, 2015) list three challenges of face this field. The first one is robustness. If one could propose a few unifying principles that provide robust assessments or comparisons across countries and time, then this field can be benefit from it a lot. The second is accuracy. Most literature in this area only generate only lower-bound estimates. How accurate these lower-bound estimates and how can we use them in the policy world are still open questions. (Kanbur & Wagstaff, 2014). The third challenge is dimensionality. There is currently no systematic method to extend fair inequality measurement by incorporating different advantage variables (wealth, health, assets, etc.).

CHAPTER 2

APPLICATION I: CHINA

2.1 Existing work

Despite the fact that the theoretical and empirical literature of incoportaing fairness into inequality research is growing fast, the number of empirical applications that explicitly apply this concept under the context of China is scarce. As for the unfair inequality in the traditional Roemer model, Zhang and Eriksson (2010) used China Health and Nutrition Survey (CHNS) dataset and find a very high share of inequality of opportunity in individual income inequality, increasing steadily from 46% in 1989 to 63% in 2006. However, the dataset CHNS is not nationally representative and some of the circumstance variables (e.g. age, survey year) included in their paper are controversial. Golley and Kong (2018) used China Family Panel Studies (CFPS) survey to measure the inequality in individual educational outcomes (measured in years of schooling) in aggregate and for each of 10 birth cohorts. They found an increasing trend of inequality of opportunity for the younger cohort and the dominant circumstance contributing to the unequal opportunities is the Hukou status at the age of 12. However, there is no current nationally representative research on Roemerian unfair inequality on the labor market outcome. As for the second level of unfair inequality proposed by the Hufe, Piechl and Kanbur (2018), no empirical work has been applied in China so far.

2.2 Data and sample selection

Our analysis is based on the China Family Panel Studies (CFPS) survey, which is a nationally representative survey of Chinese communities, families, and individuals produced by the Institute of Social Science Survey (ISSS) of Peking University. CFPS suits our research porperly for it contains individual-level biological data, economic activities, socio-economic background of the parents migration, and so on. This provides a wide range of outcome and circumstance variables which are necessary in our studies. The CFPS dataset is currently comprised of three waves (2010, 2012 and 2014). However, we only focus on the baseline 2010 data [1]. The main reason is that the CFPS dataset does not provide a long panel that is required to study unfair inequality dynamics in China.

To measure Roemerian unfair inequality using CFPS, we firstly need to estimate the functional form (1.3) $y_{it} = \delta C_i + \epsilon_{it}$. Given the available data, we select the logarithm of annual income as our outcome of interest y_{it}. This leads to excluding the individuals with no income reported or with non-positive income. We will try to include the omitted non-positive income in section 4 and analyze the difference. We then choose the following set of circumstance variables C_i: hukou status at age 3 (urban = 1); gender (female = 1); ethnic minority status (=0 if so, 1 if Han Chinese); and birth regions (north, east, center, west); parents highest education levels, which we use three dummy variables primary school, secondary and above (with no schooling as the baseline, excluded category); parents job (agricultural=1). For samples that the data for both parents education is available, we only include the higher one to represent parental education level. As for parental job, we categorize it as non-agricultural if at least one

[1] In the 2010 baseline survey, the CFPS successfully interviewed 14,960 households and 42,590 individuals, with an approximate response rate of 79%.

side of the parents is doing non-agricultural work. We also put an restriction on the age that are between 25 and 55 years old. Individuals in this age group are all adults who mostly finished their education, being active on the labor market and making life choices independent from their parents. In order to look at the birth cohort pattern, we further separates our restricted samples into 10 three-year birth cohorts.

Table 2.2.1: CFPS: All samples VS Selected Samples

VARIABLES	All sample	Selected sample
Annual Income	10712.1	15109.78
Age	39.78	42.41
Majority %	89.27%	89.28%
Hukou: Urban	15.78%	14.10%
Gender: Male	50.75%	56.95%
Birth Region:		
North	11.89%	11.99
East	34.01%	33.75%
Center	26.06%	25.14%
West	28.04%	29.13%
Parental Education		
Illiterate/Semi-literate	41.95	39.98
Primary school	25.87	31.43
Secondary	29.45	26.1
College and above	2.73	2.48
Parental Occupation		
Agricultural	64.34%	65.24%
Non-Agricultural	35.66%	34.76%
Sample size (n)	33,600	12,074

Note: All results are weighted by CFPS sample weights to be nationally representative

After we have applied the above sample restrictions on the original data set, Column 1 and 2 of Table 2.2.1 shows the statistical analysis of all samples and and after-selection samples. The after-selection samples do not statistically differ from the original samples, except for annual income and age. This is not surprising since the non-positive incomes are excluded when we take the log form of the income in our analysis in the Roemerian way of measuring unfair inequality. Therefore the selected samples have a higher level of income in average. Since we also restrict the age from 25-55, we excluded the very young population in our analysis.

For the HKP method, we also need similar circumstances to partition the populations into types. However, we need to do some adjustments instead of using the exact same set of circumstance as do in Roemerian method. Our goal is not to have observations fall short on each type to make the result reliable. We first take out ethnicity in our circumstance set, for the minority population is too small (10.72%) compare to the majority (89.28%). After this, we need to partition less finely within a circumstance variable. Under five circumstance variables included – biological sex, hukou registration, birth region, parental education and occupation – we further merged the sub-categories under birth region and parental education. Under birth region, we only have two categories of whether the person was born in inland or at the coastal area. Parental education variable now only reflects the two groups of people whose parents either have no education at all or they were ever educated. Then the populations is partitioned into 2 *2 *2 * 2* 2 = 32 non-overlapping circumstance types. Further more, we only keep those types for which we have a at least 25 observations in order to make the results more reliable and representative. The HKP measurement also requires the setting the line of basic need, which is resented by a poverty line in our case. The poverty line we used here is at annual income less than 2300RMB/person, using the 2010 price. This is the third and the latest official poverty line that was announced by the Chinese government in history, which better reflects the country's economic development and rising standard of living. The detailed statistics for each of the ten cohort can be find in Appendix B Table B.1.1.

2.3 Results

2.3.1 Roemerian Measurement

The measurement results of Roemerian unfair inequality are shown in Table 2.3.1 below. Panel A shows the total income inequality measured by GE(0) for the total population and for each of the 10 birth cohort. The restricted sample has a total outcome inequality of 0.66 measured by GE(0). The absolute Inequality of opportunity index IOA=0.16 and the relative Inequality of opportunity index IOR=0.24. This means that Roemerian unfair inequality takes up around 24% of the overall inequality. Compare to the work of Brunori et.al (2013) which measures the cross-sectional IOR around the world using a similar way, China is at the club of high inequality of opportunity, being far away from the lowest (Norway 2%) and the coming close to the highest (Guatemala 34%).

Table 2.3.1: Roemerian unfair inequality in China

	All	55-57	58-60	61-63	64-66	67-69	70-72	73-75	77-79	80-82	83-85
Panel A: Outcome inequality											
GE(0)	0.66	0.85	0.64	0.63	0.66	0.6	0.64	0.77	0.62	0.64	0.63
Panel B: Roemerian unfair inequality											
IOA	0.16	0.13	0.13	0.13	0.15	0.12	0.17	0.2	0.16	0.2	0.16
IOR	0.24	0.15	0.2	0.2	0.23	0.2	0.26	0.26	0.26	0.31	0.25

Figure 2.3.1 plots out the birth cohort pattern of Roemerian unfair inequality. The blue line shows the overall inequality, the red line shows IOA and the

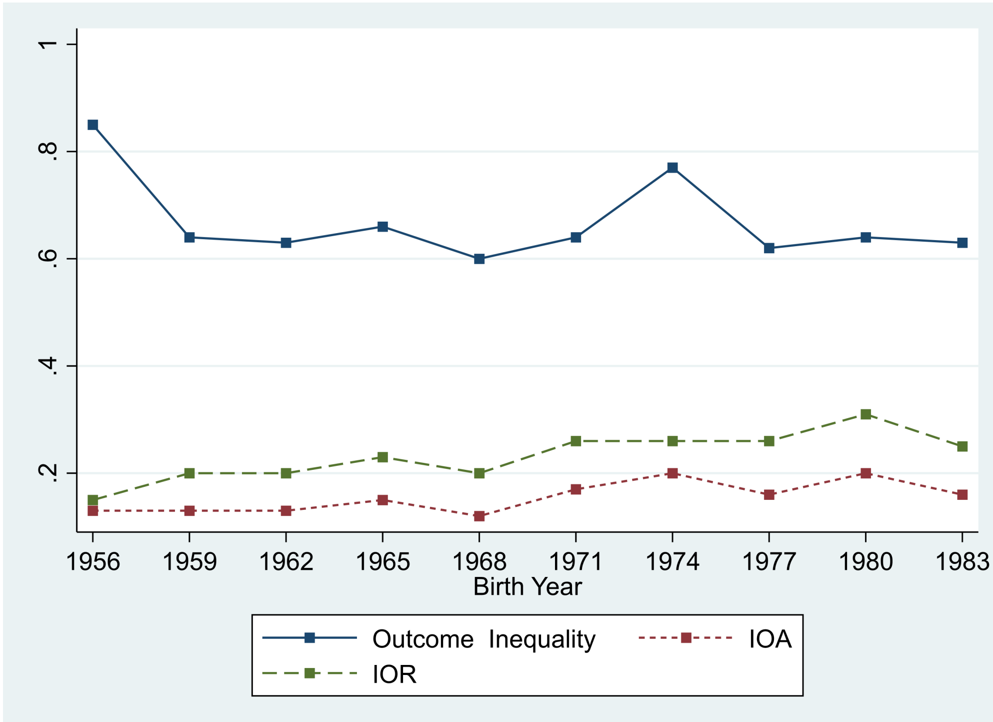

Figure 2.3.1: Roemerian Unfair inequality in China: Birth cohort pattern

green line show the IOR. We can see that the unfair inequality for the later cohort increases both in the absolute term and the relative term. People who are born after the 70s experience a higher unfair inequality compare to the national average.

Table 2.3.2: CFPS: Estimation of the log annual incomes on circumstances

VARIABLES	(1) All	(2) 55-57	(3) 58-60	(4) 61-63	(5) 64-66	(6) 67-69	(7) 70-72	(8) 73-75	(9) 77-79	(10) 80-82	(11) 83-85
Han Chinese (YES = 1)	0.24***	-0.02	0.53**	0.05	0.14	0.09	0.23**	0.31**	0.50***	0.23	0.35*
Hukou at age 3 (Urban = 1)	0.44***	0.38**	0.28	0.56***	0.60***	0.38***	0.58***	0.47***	0.18	0.37**	0.48***
Male (YES = 1)	0.80***	0.78***	0.81***	0.84***	0.83***	0.77***	0.82***	0.86***	0.80***	0.84***	0.70***
Birth region: Baseline North											
East=1	0.41***	0.48**	0.36	0.32**	0.24	0.22*	0.40**	0.62***	0.34*	0.58***	0.59***
Mid=1	0.22***	0.17	-0.01	0.16	0.08	0.19	0.21	0.29	0.21	0.64***	0.24
West=1	0.09	0.01	-0.01	-0.03	0.04	-0.03	0.07	0.15	0.02	0.52***	0.15
Parental Education: Baseline Illiterate/Semi-literate											
Primary=1	0.24***	0.31***	0.32**	0.02	0.14	0.21**	0.32***	0.37***	0	0.39***	0.46***
Secondary=1	0.32***	0.49***	-0.02	0.2	0.32***	0.12	0.37***	0.64***	0.15	0.47***	0.38***
College and above=1	0.66***	-0.37	0.13	0.57**	0.51***	0.68***	0.96***	0.53**	0.37	0.90***	0.89***
Parental job: Baseline non-agricultural											
Agricultural=1	-0.24***	-0.2	-0.17	-0.15	-0.27***	-0.25***	-0.20**	-0.22**	-0.44***	-0.47***	-0.08
Birth cohort	Yes	-	-	-	-	-	-	-	-	-	-
Constant	7.63***	7.84***	7.49***	8.25***	8.21***	8.33***	7.95***	7.70***	8.29***	7.79***	7.77***
Observations	12,074	1,150	954	1,271	1,515	1,496	1,530	1,215	1,022	951	976
R-squared	0.164	0.124	0.122	0.141	0.148	0.149	0.166	0.185	0.177	0.196	0.166

Note: All figures are weighted by CFPS sample weights to be nationally representative.
Further details on birth cohorts and robust standard errors available on request.
*** p¡0.01, ** p¡0.05, * p¡0.1

This parametric model also allows us to analyze the impact of the individual components in the circumstance set C. The detailed regression results for the functional from discussed in 3.1 are reported in Table 2.3.2. Column (1) shows the result for the overall population, with 9 cohort dummy variables included. Column (2) - (11) shows the regression results for 10 birth cohorts. Overall, we can see that being Han, having an urban Hukou, being male, being born in east area and having parents with high education or a non-agricultural occupation all have a positive influence on the income of the offspring. Of all the circumstance variables, gender the largest marginal effect. Being male gives rise to an 80% higher income of the offspring. One or both of the parents have a college degree or above also shows pronounced marginal effect, adding 66% to the childrens income. Having an urban Hukou at age of three contributes an additional 44% rise in the childrens income. Although the impact of parental occupation and ethnicity is significant, their contributions are relatively low compared to other circumstances. As for the results of 10 birth cohort, we can see that throughout column (2) to (11), gender remains as a significant circumstance and maintains its strong marginal effect. However, the magnitude of shows a decreasing trend for the younger generation. As for other circumstances, their effects on off-springs individual income varies across cohorts.

To better quantify the changes of these circumstances contribution, we can repeat the robust OLS estimation by dropping the circumstance of interest (e.g. gender) one by one. Next, we obtain a new predicted individual income $\tilde{y}_{nogender}$, then calculate the inequality for $\tilde{y}_{nogender}$. The ratio $\frac{I(\tilde{y})-I(\tilde{y}_{nogender})}{I(\tilde{y})}$ shows how much the circumstance factor gender contributes to inequality of opportunity. Fig 2.3.2 shows the contributions of the individual circumstance to the overall Roemerian unfair inequality over time. The detailed statistics can be find in Ap-

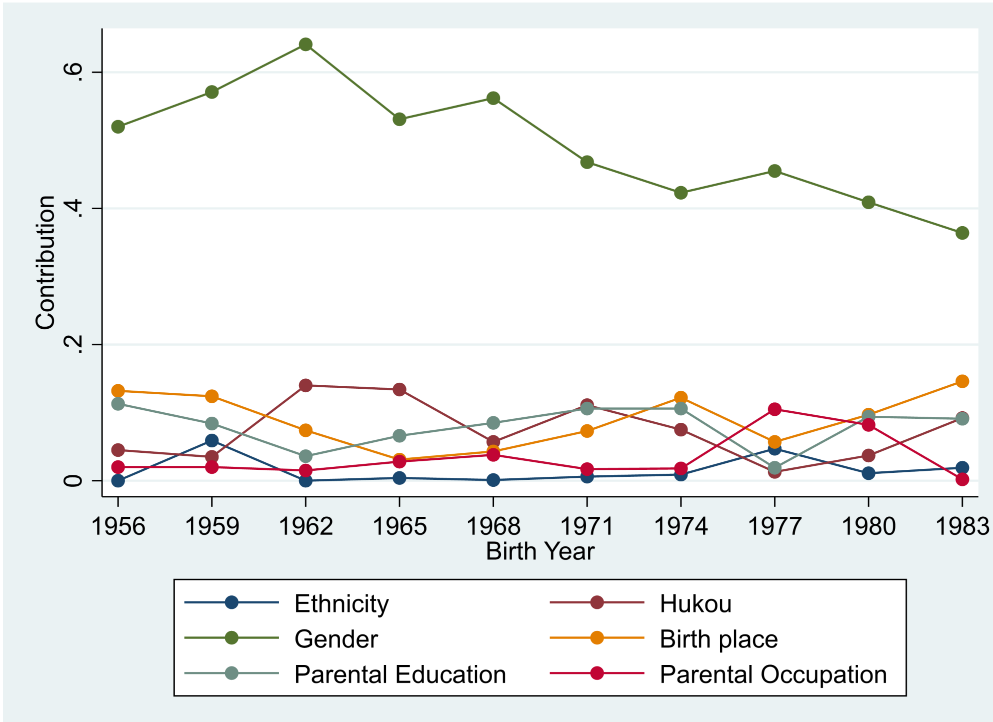

Figure 2.3.2: Contribution of different circumstances

pendix E Table E.1. The first thing worth noticing is that the gender contributes the most to the unfair inequality. On average gender accounts for 49% of IOA. Another two important variables are birthplace and parents' education, which contribute on average 9% and 8% respectively.

2.3.2 HKP Measurement

Results for HKP unfair inequality are shown in Table 2.3.3 in Panel C, which measures the aggregate divergence between the empirical distribution Y and the reference distribution Y^r. Divergences $D(Y||Y^r)$ are aggregated based on mean log deviation. The total HKP unfair inequality is 0.18, which now takes up 27% of the total inequality. This upward correction is expected due to the fact that we include the other principle FfP. Although the magnitude of the correction is

Table 2.3.3: HKP unfair inequality in China

Birth year	All	55-57	58-60	61-63	64-66	67-69	70-72	73-75	77-79	80-82	83-85
					Panel A: Outcome inequality						
GE(0)	0.66	0.85	0.64	0.63	0.66	0.6	0.64	0.77	0.62	0.64	0.63
					Panel C: HKP unfair inequality						
Total HKP	0.18	0.2	0.21	0.2	0.22	0.17	0.19	0.2	0.17	0.21	0.18
Share of HKP	0.27	0.24	0.33	0.32	0.33	0.28	0.3	0.26	0.27	0.33	0.29
UBEOp	0.08	0.07	0.06	0.11	0.1	0.11	0.09	0.14	0.11	0.11	0.09
LBFfP	0.1	0.13	0.15	0.09	0.12	0.06	0.1	0.08	0.07	0.09	0.09

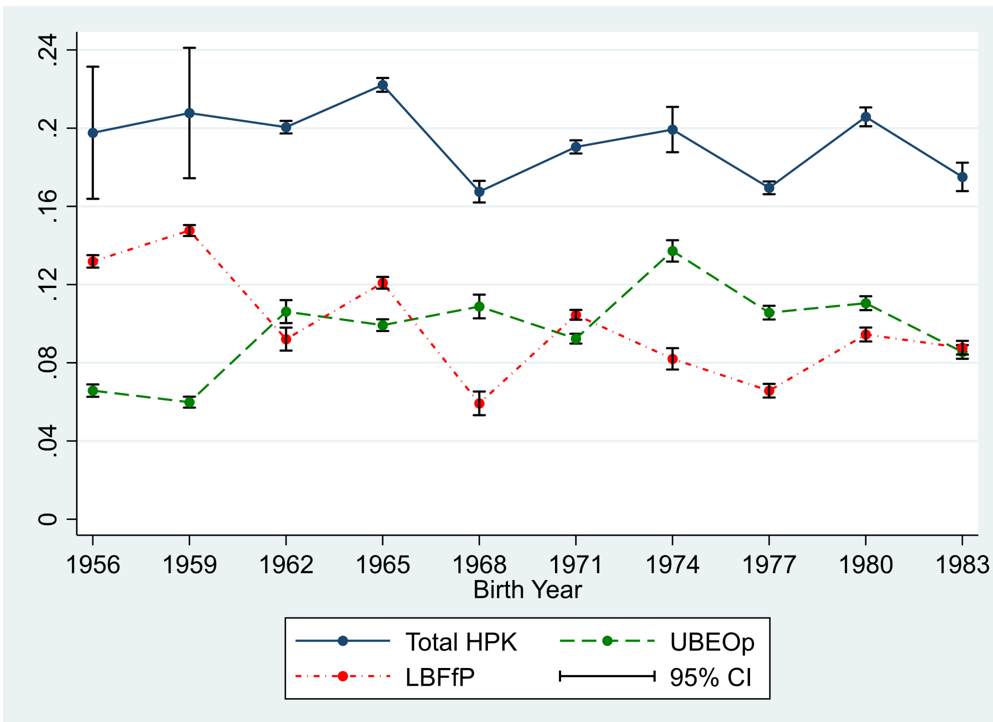

Figure 2.3.3: CFPS: Contribution of different circumstances

small, it maintains for each of the ten birth cohort. In Figure 2.3.3, we plot out the HKP unfair inequality across cohort using the blue line. The black vertical lines represent 95% confidence intervals, which are calculated based on a bootstrap procedure with 400 draws. Unlike the Roemerian unfair inequality in Fig 1, we find that the HKP unfair inequality experience a drop in the late 60s and then remains relatively steady around 0.18 for later cohorts.

To better explain this stability, we are certainly interested in discovering how this unfair inequality is obtained by violations of EOp or FfP separately. According to Hufe, Piechl and Kanbur (2018), the construction of the norm distribution implies that the unfair inequality that violates either FfP or EOp, are partly overlapping. Therefore. it is not achievable to completely separate them. For example, re-scaling all incomes in a type to realize EOp might also lift a portion of people out of poverty. Similarly, moving all the individuals below the poverty line up to the poverty line to realize FfP might make the poors income closer to the population mean. Therefore, how much unfair inequality attributable to either of the principle alone gives an the upper bound share of HKP unfair inequality. The respective lower bound can be measured as how much unfair inequality is left uf we only follow the other principle. Thus we can decompose HKP unfair inequality as $D(Y\|Y^r_{EOp+FfP}) = D(Y\|Y^r_{EOp+FfP}) - D(Y\|Y^r_{EOp}) + D(Y\|Y^r_{EOp})$ where the first two terms represent lower bound of Ffp and the last term represents the upper bound of EOp. The detailed construction of $Y^r_{EOp+FfP}$, Y^r_{EOp}, and Y^r_{FfP} are given in Appendix A. The results for the upper bound of unfair inequality that violates only EOp (UBEOp) and the lower bound of unfair inequality that violates only FfP (LBFfP) are shown in the last two rows of Panel C in Table 2.3.3.

In Fig 2.3.3, the red line shows LBFfP and the green line shows UBEOp. We first notice an increasing trend of UBEOp, which is in line with our findings for the Romerian unfair inequality in Fig 2.3.1. However, LBFfP decreases for the later cohorts. This two opposite trend of LBFfP and UBEOp explains the relative stability of HKP unfair inequality after the 60s. Our results reflect that the younger generations face less unfair inequality to freedom from poverty but facing a bigger problem of a reduction in equal opportunity. The older generations have relatively equal opportunity but have more trouble meeting the basic needs.

2.3.3 Comparison

Table 2.3.4 combines the unfair inequality measurement result for both the Roemerian method and the HKP method. We see that overall, HKP method yields a higher unfair inequality when we compare IOA with total HKP. The exceeding part is expected since HKP include the principle of FfP, creating a longer distance for some poor to reach the poverty line instead of just the mean income of its type. However, both of the methods still yields a lower bound of the unfair inequality of the society if we couldnt include all the proper circumstances that individuals face.

Another interesting difference to notice is that the upper bound of HKP unfair inequality violating EOp (UBEOp) is lower compared to Roemerian absolute inequality of opportunity IOA. This is due to the non-parametric nature of HKP method. In order to obtain enough observation for each type, we have to cut down the number of circumstance variables or make the existing ones more coarsely defined. This adjustment reduces the explanatory power of the circumstance. The parametric Roemer model doesnt have this problem of keeping a certain number of observations in a type cell. But it has another problem of overlooking the interplay of different circumstance. In our case, the loss of explanatory power due to fewer circumstance included is clearly larger than the loss due to ignoring the interplay between the existing circumstances. If we want to pick a magnitude of inequality of opportunity, then the Roemerian

Table 2.3.4: CFPS: Comparison between Roemerian and HKP unfair inequality

Birth Year	All	55-57	58-60	61-63	64-66	67-69	70-72	73-75	77-79	80-82	83-85
Panel A: Outcome inequality											
GE(0)	0.66	0.85	0.64	0.63	0.66	0.6	0.64	0.77	0.62	0.64	0.63
Panel B: Roemerian unfair inequality											
IOA	0.16	0.13	0.13	0.13	0.15	0.12	0.17	0.2	0.16	0.2	0.16
IOR	0.24	0.15	0.2	0.2	0.23	0.2	0.26	0.26	0.26	0.31	0.25
Panel C: HKP unfair inequality											
Total HKP	0.18	0.2	0.21	0.2	0.22	0.17	0.19	0.2	0.17	0.21	0.18
Share of HKP	0.27	0.24	0.33	0.32	0.33	0.28	0.3	0.26	0.27	0.33	0.29
UBEOp	0.08	0.07	0.06	0.11	0.1	0.11	0.09	0.14	0.11	0.11	0.09
LBFfP	0.1	0.13	0.15	0.09	0.12	0.06	0.1	0.08	0.07	0.09	0.09

result might be a better one in our case. Yet this difference doesnt change the trend of unfair inequality due to unequal opportunity that we observe. For both of the measurement, we observe an increasing trend of unfair inequality stemming from unequal opportunity.

Although we can cross-validate the trend of EOp using the two methods, the features of unfair inequality violating FfP can only be obtained using HKP method. From Table 2.3.4 Panel C, we find that unfair inequality to FfP takes nearly 56% of the total HKP unfair inequality. This is not a small size at all. Bringing in the HKP method certainly generate a new dimension when viewing unfair inequality and suggests a shift of the recent focus of attention from the uppermost parts of the income distribution to the lower percentiles.

In all, the Romerian method and the HKP method can be a good complement to each other when we want to get a fuller picture of the how much of the inequality is unfair and how it changes.

2.4 Including Zero Incomes

Most literature analyzing unfair inequality in individual income omit zero income and/or negative income. This usually happens under these two procedures: 1) The log form of the income is used to stabilize the variation in the parametric models. 2) The prevalent use of general entropy class as an inequality measurement index. This is certainly unsatisfactory due to the fact that non-positive income is quite common among a large amount of income surveys. For example, about 400 of the 700 income surveys used in the World Bank's PovcalNet for global poverty and inequality measurement have non-positive values for household income (Ravallion, 2017). By excluding all the non-positive income in the analysis would certainly underestimate the degree of poverty and inequality.

To include non-positive income in our measurement, we utilize a concave log-like transformation proposed by Ravallion (2017). The transformation is a hybrid of hyperbolic sine and its inverse:

$$h\left(y;a\right) \equiv I sinh(\alpha y) + (1-I)\, sinh^{-1}\left(\alpha y\right) - ln2\alpha \;\; (\alpha > 0) \tag{2.1}$$

where y represents income, and I takes the value unity if $y \leq 0$ and zero otherwise. In other words, this method uses the ordinary hyperbolic sine transformation for negative y, and only use the inverse function for positive y. Ravallion (2017) also provides a modified mean log deviation in which lny is replaced by $h\left(y;a\right)$. The formula for this modified version is:

$$H\left(\alpha\right) \equiv h\left(\overline{y};\alpha\right) - \frac{1}{n}\sum_{i=1}^{n} h\left(y_i;\alpha\right) \tag{2.2}$$

where $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. By comparing $H\left(\alpha\right)$ to MLD, Ravallion found that $H\left(\alpha\right)$is substantially higher. This suggests that how one deals with negative net in-

Table 2.4.1: Unfair inequality (with 0 income included)

	All	55-57	58-60	61-63	64-66	67-69	70-72	73-75	77-79	80-82	83-85
Panel A: Outcome inequality											
GE(0)	1.99	2.65	2.1	2.19	1.96	1.78	1.75	2.09	1.9	1.98	1.85
Panel B: Roemerian unfair inequality											
IOA	0.93	0.99	0.71	0.8	0.8	0.92	0.98	0.92	0.99	1.02	0.96
IOR	0.47	0.37	0.34	0.36	0.41	0.52	0.56	0.44	0.52	0.52	0.52
Panel C: HKP unfair inequality											
Total HKP	1.49	2.04	1.6	1.62	1.54	1.43	1.33	1.51	1.41	1.44	1.24
Share of HKP	0.75	0.77	0.76	0.74	0.79	0.8	0.76	0.72	0.74	0.73	0.67
UBEOp	0.14	0.12	0.09	0.17	0.14	0.17	0.14	0.19	0.16	0.18	0.13
LBFfP	1.37	1.92	1.5	1.46	1.4	1.26	1.19	1.35	1.25	1.27	1.11

come can make a big difference to measures of inequality using log or log-like transformations. For large values of α ,the hyperbolic sine transformation yields very large negative numbers at the negative extremes. Values of $\alpha < 1$, especially $\alpha = 0.5$, are probably a sensible choice if one is measuring inequality using the h-transformation. Hence, following Ravallion's (2017) idea, we use the concave log-like transformation ($\alpha = 0.5$) to deal with the 2560 individual with zero income in our dataset. Appendix D Figure D.1.1 compares the kernel density between the income that omits zero and the incomes that includes 0 after transformation. Appendix D Table D.2.1 shows the regression result in the parametric model when we use the after-transformation income.

Using this new income, we re-esitmate the two levels of unfair inequality in Section 2.3. The results are reported in Table 2.4.1. We can see when we include 0 income, the overall inequality measured by GE(0), Roemerian unfair inequality and HKP unfair inequality all become significantly higher. In panel B, Roemerian unfair inequality takes up to 47% percentage of the total outcome inequality in average. This result is approximately twice as the IOR we measured in Table 2 Panel B. The extent of this upsize correction also exists for 10 birth cohorts.

This is within our expectation since including the non-positive incomes creates more variations of the mean incomes across types. In Table 2.4.1 panel C, the results for HKP unfair inequality is 1.49, which is more than 8 times higher than the total HKP we get in Table 2.3.4 Panel C, taking up nearly 75% of the overall inequality. This large increase in the size of HKP unfair inequality also exists for each cohort. Clearly, the upward correction is much more distinguishable for the HKP method. Since the HKP method incorporates the principle of FfP, a large flow of extremely low income (2560 zero income) could certainly surge up the unfair inequality due to the violating of FfP. Indeed, after the decomposition, we find the increase of LBFfP obviously dominates the increase in UBEOp. Fig 4 plots out the birth cohort pattern for the Roemerian unfair inequality results in Table 6 Panel B when we include 0 income. Compare Fig 2.4.1 to Fig 2.3.1, although the values for each cohort are different, the increasing trend of Romerian unfair inequality both in the absolute term and the relative term still remains.

Figure 2.4.2 above shows the contribution of each circumstance after we include the non-positive incomes. Compare to Figure 2.3.2, gender becomes a more dominating circumstance, explaining 73% of the overall unfair inequality. Whats more, the decreasing impact of gender in Fig 2.3.2 now doesnt exist in Figure 2.4.2. Starting from the cohort who were born around 1965, gender has an increasing impact on creating more unfair inequality. This implies that genders impact among the poor in much more strong. The overall significance of Hukou and Ethnicity decrease.

Fig 2.4.3 plots out the birth cohort pattern for HKP unfair inequality results in Table 6 Panel C. Comparing Fig 2.4.3 to Fig 2.3.3, we first notice that the over-

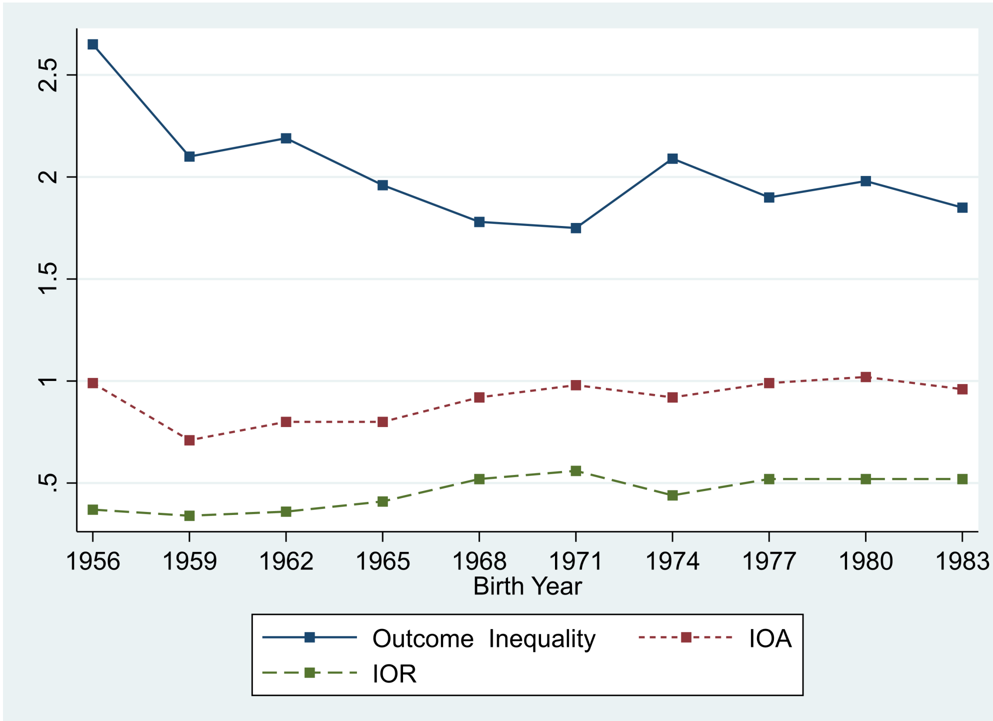

Figure 2.4.1: Roemerian Unfair inequality: Birth cohort pattern (With 0 income)

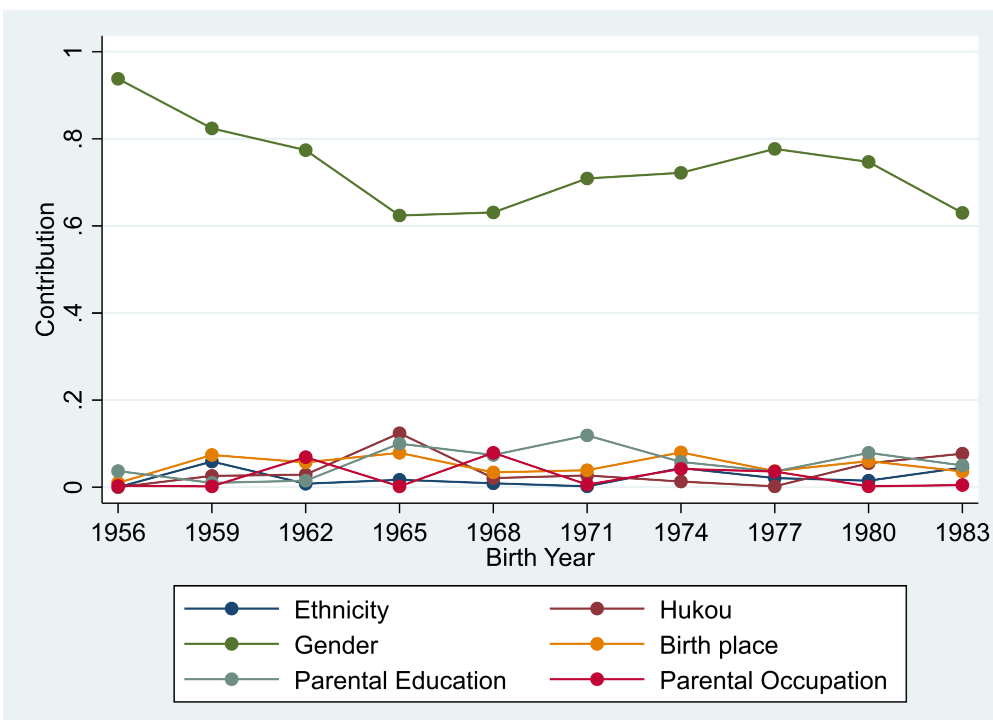

Figure 2.4.2: Contribution of different circumstances (With 0 income)

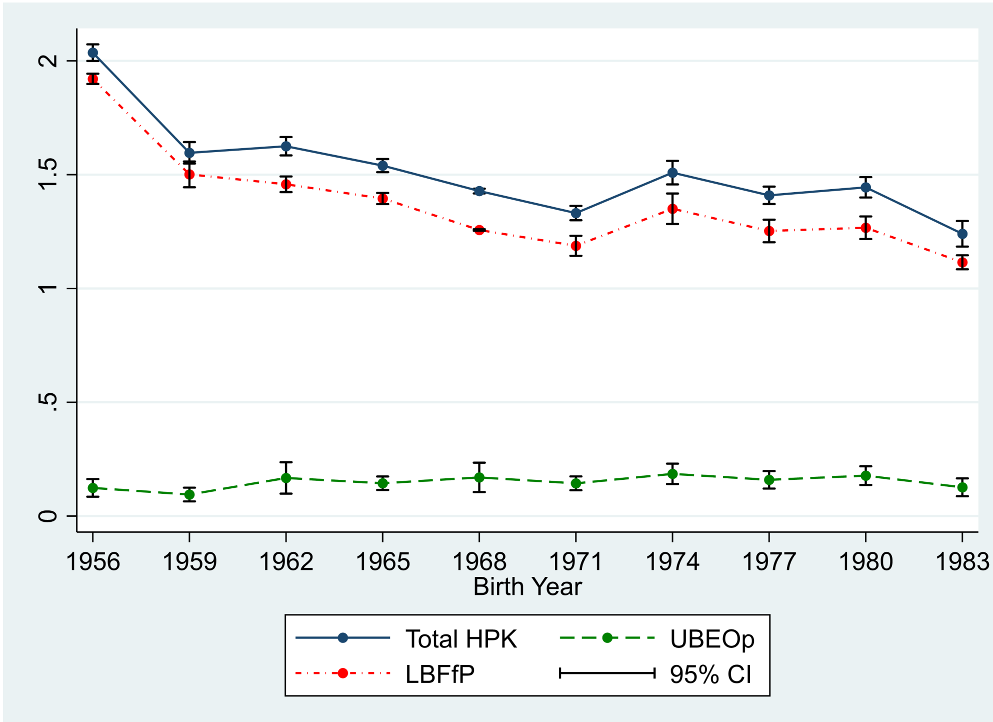

Figure 2.4.3: HKP Unfair inequality: Birth cohort pattern(With 0 income)

all trend changes. HKP unfair inequality now decreases for the younger cohorts. This is due to the decreasing trend of LBFfP dominates the change of HKP unfair inequality over birth cohorts. As for the trend of UBEOp, it remains a slightly insignificant increase. However, the ratio of UBEOp significantly decreases.

As a conclusion, including the non-positive income in our analysis significantly increases total our unfair inequality and its share in the overall inequality. The role of gender contributing to unequal opportunity gets reinforced. Yet this doesnt change the birth cohort pattern of increasing unfair inequality to EOp and decreasing unfair inequality to FfP. Although this transformation is not widely used nor does its precision gets tested out by a lot of researchers, our paper provides the first attempt at incorporating 0 income in the unfair inequality analysis. We dont intend to argue that the results are ideal. Instead, showing this huge size of the upward correction should be concerned for future research in the unfair inequality measurement. One of the critics of the current unfair inequality measurement is that its lower bound nature often leads to a very low estimates (Kanbur & Wagstaff, 2014). The huge size of upward correction in our results might provide a new perspective to address this concern. The estimate of unfair inequality is low might simply because we left out the ultra-poor in our analysis.

2.5 Conclusion

This paper presents the first nationally representative analysis of Chinas unfair inequality in individual income using data from the China Family Panel Studies Survey. Unfair inequality is measured in two ways. We first focus on the unfair inequality that violates the normative principle of Equal Opportunity (EOp). Following the conceptual contribution of Roemer (2006), we use an ex-ante parametric model to examine to what extent individuals' incomes inequality is due to circumstance factors, that is, factors beyond their control. This method allows us to calculate the contribution of each circumstance by comparing the outcome when we take each circumstance out of the model one by one. However this single definition of fairness is not enough. Another essential normative principle needed is satisfying the basic need. Therefore, we take a step further to measure the inequality that violates both EOp and Freedom from Poverty FfP based on Hufe, Piechl and Kanbur (2018). This method allows us to decompose the unfair inequality into its EOp and its FfP components.

The first measure of unfair inequality (we also call it Roemerian unfair inequality) indicates that 24% of the total outcome inequality is unfair. We further separate the samples into 10 cohorts and find that this Roemerian unfair inequality increases for the younger cohort. Among all the circumstance variables we include, gender significantly dominates all others by explaining 55% of the unfair inequality. Our second measure of unfair inequality, which we called HKP unfair inequality, takes around 27% of the unfair inequality. By further decomposing HKP unfair inequality into its EOp and its FfP components, we find 45% stems from violating EOp and 55% stems from violating FfP. Over different cohorts, HKP unfair inequality remains relatively stable. This is due to the combination of an increasing trend of unfair inequality to EOp and a decreasing trend of unfair inequality to FfP.

Our results reflect a high degree of unfair inequality in China. Inequality of opportunity is substantially correlated with gender inequality. As for different cohorts, the younger generations face less unfair inequality to freedom from poverty but facing a bigger problem of a reduction in equal opportunity. The older generations have relatively equal opportunity but have more trouble meeting the basic needs. Policy implications are inferred. Given that gender is among the most influential of circumstances, policy interventions aimed at reducing the gender pay gap should be considered in order to create more equal opportunity. Basic social security networks, especially for the older generation, should also be considered.

This paper also points to an important technical question in the unfair inequality measurement literature. The prevalent modeling convention is to omit the non-positive income when the log form of income and/or mean log deviation is used in the procedure. This would clearly underestimate the level of unfair inequality. We used a concave log-like transformation proposed by

Ravallion (2017) to see how including a large amount of non-positive incomes would change our results. The finding is quite interesting. After we include non-positive income in our analysis, we find a significantly higher amount of unfair inequality. Roemerian unfair inequalitys share of total inequality now becomes 47%, which is 2 times larger compare to the previous result. The HKP unfair inequality now takes up 75% of the total inequality, of which the amount turns nearly three times larger. This substantial increase is present for all 10 cohort as well. In conclusion, we hope our finding would pique the interest of scholars in the unfair inequality measurement area. The inclusion of the ultra-poor is essential and they would make a big difference. Future research on how to properly include the non-positive income or missing income is necessary. The controversy with respect to the usefulness of the low estimates in this area may be due to omitting those zero, negative or missing incomes.

CHAPTER 3

UNFAIR INEQUALITY IN SOUTH AFRICA

3.1 Existing work

Research on the topic of unfair inequality in South Africa is rare, especially on the labor market. In 2012, World Bank has a report that focuses on inequality of opportunity in South Africa. The first part of report discussed the opportunities available to children using Human Opportunity Index (HOI). The second part focus on inequality of opportunities in employment. According the report, date from Quarterly Labour Force Surveys (2008q1 to 2012q1) is used. Because of data limitations, only employment is included. circumstances (gender, ethnicity, and location) and education and age of the workers are considered as the determinants of his or her employment status. The results indicate more than half the inequality of opportunity stems from education and age. The direct contribution of circumstances has fallen or remained unchanged in the past four years. Moreover, the contribution of education has increased, especially for employment in the formal sector that is non-agriculture.

Another paper that focus on unfair inequality on labor market in South Afirca is written by Piraino (2015), who used NIDS (National Income Dynamics Study) from 2008 to 2012 (3 waves) to study the intergenerational earnings mobility and inequality of opportunity in South Africa. In his paper, the outcome variable is individual gross income; the circumstances variables are race, birth province, fathers education and occupation. The sample is restricted to men between 20 and 44 years. Using both parametric and non-parametric method, the results show that the IOp ranges from 0.171-0.241. A comparison of these results to other countries (Brunori, Ferreira, & Peragine, 2013) reveals that South Africa is at the upper end of the international distribution of opportunity inequality, even with the limited set of circumstances included in the analysis. A most recent working paper (Hufe, Peichl, & Daniel, 2019) estimates the upper and lower bound of the share of inequality of opportunity in the overall inequality, which ranges from 3.24% - 73.94%.

3.2 Data and Sample selection

This empirical study utilized the National Income Dynamics Study (NIDS). This is the first national panel data set implemented by the South Africa Labor and Development Research Unit (SALDRU) at the University of Cape Town. NIDS is the ideal dataset with which to analyze unfair inequality in South Africa, because it contains detailed individual level data like labor income, socio-economic background of the family, demographic characteristics, etc. The NIDS dataset is currently comprised of five waves (collected in 2008, 2010-2011, 2012, 2014-2015, 2017). Simialr to the case of CFPS, the panel itself is not long enough to analyze the dynamic change of the unfair inequality in South Africa. Therefore, we pooled and reweighted the samples from five wave, in order to get more observations and at the same time keeping our result remain nationally representative. For individuals appearing in a single wave of survey, the cross-sectional post-stratified weight is used. The post stratification weights are calculated as the probability of an individual being included in the NIDS sample was calculated using a two-stage cluster survey design. For individuals whose information is available in more than one wave, the relevant panel weight is applied. The panel weights provided by NIDS can be used to correct for attrition.[1]

For outcome variables, we use individual gross income. NIDS doesnt have individual expenditure available, hence this largely excludes the population that relies on non-labor income - a main source of earnings for a majority of the population. By using per capita household income or per capita household expenditure, a significant portion of the South African population can be accounted for in analyses of unfair inequality(Leibbrandt, Woolard, McEwen, & Koep, 2010). However, the per capita household income are not a good representative of a personal advantage, since it ties up so closely to the household condition. Therefore, we still choose to use individual gross income.

For the circumstance variables, we include gender, race, parental education and occupation. We further proxy the occupational status of parents by grouping them in three levels: never worked, elementary occupations, or non-elementary positions. Parents education level is also separated into four classes: 1) no schooling; 2) 0-7 grades; 3) 8-11 grades; 4) Metrics and above. For samples that both parents information is available, we take the highest education and highest class of job among the two as the parental education and occupation. The classification detail is given in Appendix B Table A.3.We only retain information on the parent of highest occupational status.

We first keep the observations with individual gross income and above circumstance variables available in our sample. The sample is further restricted to adults aged between 25 and 55 years old. we choose this age range for the same reason as in CFPS data - people in this age range mostly finished education and are active on the job market. In Piraino (2015), the sample is restricted

[1]For example, the wave1-to-wave2 weight if appearing in the first two waves, the wave1-to-wave3 weight if appearing in all three waves and so on.

to men between 20 and 44 years of age for purpose of keep a good sample size with fathers information available. He only used male sample to keep it consistent with the intergenerational mobility analysis in the same paper using father and son data. Our analysis not only expands the age range but also includes females. After we have applied the above sample restrictions on the original data set, Column 1 and 2 of Table 3.2.1 compare the full and selected samples. The descriptive statistics show that the samples are not similar. The selected samples have higher gross income and age in average. It is also more prevalent for the selected samples to have parents working as in non-elementary job and have higher levels of education. It is not surprising that the restricted sample statistically differed from the full sample on these characteristics. Our sample first excludes the extreme young and old and thus cutting off many low or 0 income. A large amount of the women population is unemployed in South Africa. The employed individuals who reported zero labor market income are excluded when we take the log form of the income in our analysis in the Roemerian way of measuring unfair inequality.

For the HKP method, we also need similar circumstances to partition the populations into types. However, we need to do some adjustments instead of using the exact same set of circumstance as do in Roemerian method. Our goal is not to have observations fall short on each type to make the result reliable. Since we dont want observations to fall short on each type, we need to make the partition less finely within a circumstance variable. Under four circumstance variables included – biological sex, race, parental education and job – we further merged the sub-categories under race birthplace, parental education and job. For racial indicator, we separate the population into African and non-African. We separate parental education into 2 types: no schooling and with

Table 3.2.1: NIDS: All samples VS Selected Samples

VARIABLES	Overall	Selected
Gross Income(Rand/year)	21177	27524
Age	29	39
Female %	51.56	45.56
Race %		
African	79.3	75.95
Colored	9.11	8.74
Asian/Indian	2.58	2.65
White	9.01	12.66
Parental job: %		
Never worked	40.52	29.85
Elementary	24.7	22.26
Non-elementary	34.78	47.89
Parental education: %		
No Schooling	48.39	33.62
0-7 grade	19.59	19.72
8-11 grade	19.91	27
Matric+	12.11	19.66
Observations	46,449	9,076

Note: All results are weighted by NIDS sample weights to be nationally representative

education. For parental occupation, instead of using five classes, we reduce the class number into 2, where the first class is never worked plus elementary work. The second class is non-elementary. Then we can partition the population into 2 *2 *2 *2 = 16 types. We don't want small number of observations to fall in a type and thus lose the precision our result. Hence, we only keep the types with a minimum of 20 observations.

The poverty line we used there is the national poverty line announced by South Africa government. The national poverty lines were constructed using the cost-of-basic-needs approach which links welfare to the consumption of goods and services. The lines contain both food and non-food components of household consumption expenditure. Our baseline analysis will use food poverty line, which is 515 rands per person per month in 2017s price. This data is provided by Statistics South Africa. We multiply this poverty line by 12 and compare it with the inflation adjusted annual income. The detailed statistics for each of the seven cohort can be find in Append C table C.2.1.

3.3 Results

3.3.1 Roemerian measurement

The measurement results of Roemerian unfair inequality are shown in Table 3.3.1 below. Panel A shows the total income inequality measured by GE(0) for each of the survey year. On average, the restricted sample has a total outcome inequality of 1.49 measured by GE(0). The absolute Inequality of opportunity index IOA=0.24 and the relative Inequality of opportunity index IOR=15.90%. This means that from 2008 to 2017, Roemerian unfair inequality takes up around 15.8% of the overall inequality in average. Compare to the work of Brunori et.al (2013) which measures the cross-sectional IOR around the world using a similar way, South Africa is at the club of high inequality of opportunity(highest is Guatemala 34%). Piraino (2015) used NIDS 2008 to 2012 pooled samples and estimated IOA=0.17-0.24 in the gross individual income in South. The upper bound of his mwasurement is the same as our study, which is understandable given that Piraino (2012) only observed two circumstances (race and fathers education) in that particular study.

We are also interested in investigating the birth cohort pattern. After restrict our sample to the age from 25-55, we found that their birth year is from

Table 3.3.1: Roemerian unfair inequality in South Africa

	All	58-62	63-67	68-72	73-77	78-82	83-87	88-92
	Panel A: Outcome inequality							
GE(0)	1.49	1.15	1.58	1.18	0.64	0.51	0.46	1.00
	Panel B: Roemerian unfair inequality							
IOA	0.24	0.51	0.31	0.17	0.25	0.16	0.10	0.62
IOR	15.90%	44.33%	19.70%	14.54%	38.32%	30.86%	21.58%	61.53%

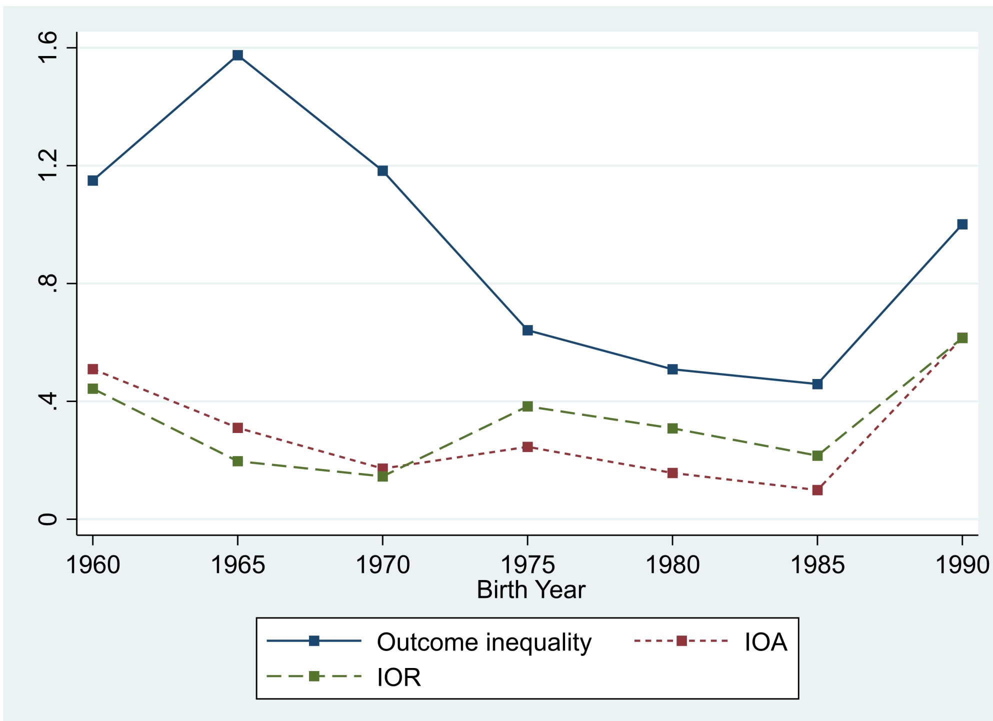

Figure 3.3.1: Roemerian Unfair inequality in South Africa: Birth cohort pattern

1958-1992. Hence we use five years as a cohort and separate our samples into 7 cohorts. Figure 3.3.1 plots out the birth cohort pattern of Roemerian unfair inequality. The blue line shows the overall inequality, the red line shows IOA and the green line show the IOR. We can see that from for the later cohort that were born from 1960 to 1985, the younger cohort has a decreasing IOA but an overall increasing IOR. For the youngest cohort that were born around the 1990s, both of their and IOA and IOR increases significantly.

This parametric model also allows us to analyze the impact of the individual components in the circumstance set C. The detailed regression results are reported in Table 3.3.2. Column (1) shows the overall result and column (2)-(8) shows the result for each of the seven cohorts. Overall, we can see three circumstances gender, race, and parental education are play significant roles in off-springs income. Parental occupation is not having a big impact on off-springs outcome. Of all the circumstance variables, race has the largest marginal effect. Overall, being Asian/Indian gives rise to a 95% higher income compare to African. For White, the rise is also high as 80%. One or both of the parents have a Metrics degree or above also shows pronounced marginal effect, adding 97% to the childrens income overall. Being female contributes a 45% drop in the individuals income. Across column (2) to (8), we notice the negative effect of being a female decreases over the years, while the negative effect of being African increases over the years. The positive effect of parents high education decrease over the years.

Table 3.3.2: NIDS: Estimation of the log annual incomes on circumstances

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	All	58-62	63-67	68-72	73-77	78-82	83-87	88-92
Female (Yes=1)	-0.45***	-0.71***	-0.47***	-0.60***	-0.51***	-0.25***	-0.13*	-0.60***
Race: Baseline African								
Coloured	0.13**	0.34**	0.15	0.18	-0.01	0.06	0.14	-0.21*
Asian/Indian	0.95***	1.49***	0.28	0.32**	0.88***	1.56***	0.89***	0.02
White	0.80***	1.43***	0.78***	0.21	0.48**	0.21	0.71***	2.10***
Parental Job: baseline Never worked								
Elementary	-0.05	-0.09	-0.18	0.13	0.08	-0.05	-0.13	-0.2
Non-elementary	0.07	-0.23	0.05	0.30***	0.06	0.12	0.06	-0.04
Parental Education Baseline: No schooling								
0-7 grade	0.27***	0.34**	0.68***	0.29***	0.09	0.01	0.25***	0.24**
8-11 grade	0.60***	0.66***	0.93***	0.65***	0.69***	0.36***	0.42***	0.37***
Matric+	0.97***	0.82***	1.21***	1.04***	1.04***	0.71***	0.79***	0.90***
Constant	9.07***	9.38***	8.89***	8.98***	8.95***	8.61***	8.25***	8.92***
Observations	9,067	983	1,292	1,497	1,585	1,433	1,499	778
R-squared	0.3	0.44	0.35	0.26	0.35	0.23	0.19	0.54

Note: All results are weighted by NIDS sample weights to be nationally representative.
Further details on birth cohorts and robust standard errors available on request.
*** p¡0.01, ** p¡0.05, * p¡0.1

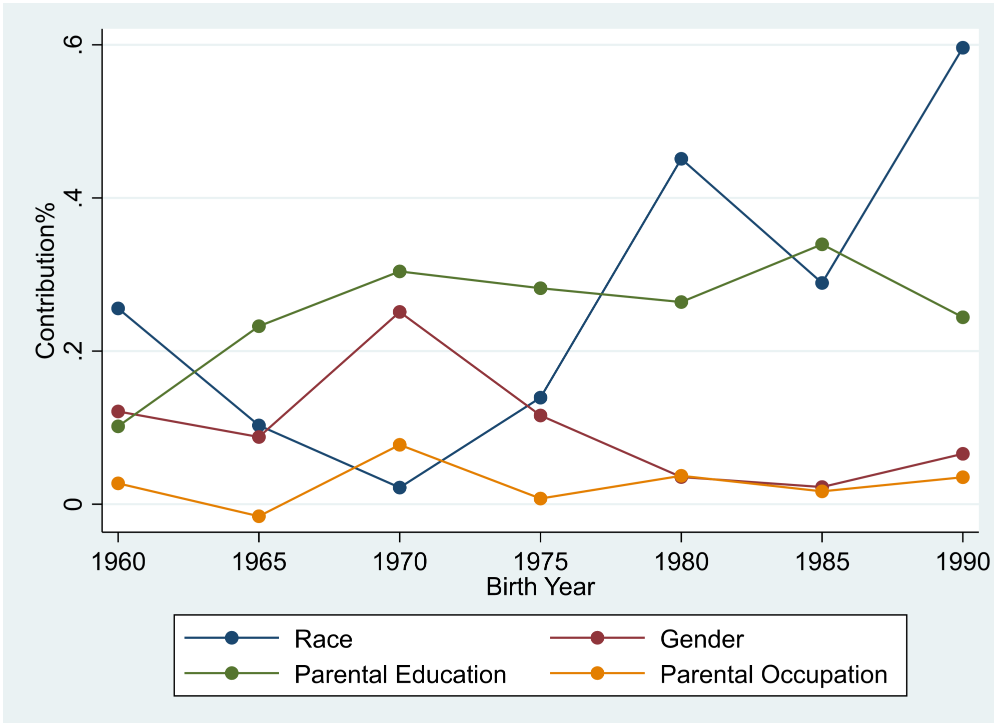

Figure 3.3.2: Contribution of different circumstances: South Africa

To better quantify the changes of these circumstances contribution, we can again carry out robust OLS estimation but now dropping the variable of interest (e.g. gender) one by one. Using the same method we discussed in 2.3.1, Figure 3.3.2 shows the contributions of the individual circumstance variables to the overall inequality of opportunity and how these have changed over time. The detailed date can be found in Appendix C Table C.2.1. The first thing worth noticing is that the race and parental education contributes the most to the unfair inequality. On average they both accounts for 23% of IOA. Another important variable is gender which contribute on average 9.44% respectively to the Roemerian unfair inequality. As for the trend, the contribution of race and parental education increases overtime while the contribution of gender decreases over time.

3.3.2 HKP measurement

Results for HKP unfair inequality are shown in Table 3.3.3 panel C, which measures the aggregate divergence between the empirical distribution Y and the reference distribution Y^r. The divergence measurement is bases on mean log deviation, making the result comparable with the Roemerian unfair inequality. The total HKP unfair inequality is 0.33 in average, with a share of 22.42% of the total inequality. This upward correction is expected due to the an increasing part of inequality that violates the FfP principle. Although the magnitude of the correction is not big, it maintains for each of the survey year. We plot out the HKP unfair inequality in Figure 3.3.3 The blue line represents the total outcome inequality. HKP unfair inequality is represented by the dash red line. As the cohorts get younger, the HKP unfair inequality experiences a drop first and then an increase.

Table 3.3.3: HKP unfair inequality in South Africa

	All	58-62	63-67	68-72	73-77	78-82	83-87	88-92
	Panel A: Outcome inequality							
GE(0)	1.49	1.15	1.58	1.18	0.64	0.51	0.46	1
	Panel C: HKP unfair inequality							
HKP	0.33	0.44	0.85	0.32	0.26	0.33	0.28	0.95
Share of HKP	22.36%	38.53%	53.79%	26.80%	39.97%	64.02%	60.03%	95.22%
UBFfP	0.28	0.22	0.21	0.21	0.19	0.32	0.29	0.25
LBEOp	0.05	0.22	0.64	0.1	0.06	0	-0.01	0.7

To better explain this trend, we are certainly interested in discovering to what extent our results are driven by violations of EOp or FfP separately. Using the same decomposition method in Section 2.3.2, we can seperate the hkp unfair inequality into the upper bound of unfair inequality that violates only EOp (UBEOp) and the lower bound of unfair inequality that violates only FfP (LBFfP). The results are shown in the last two rows of Panel C in Table 3.3.3.

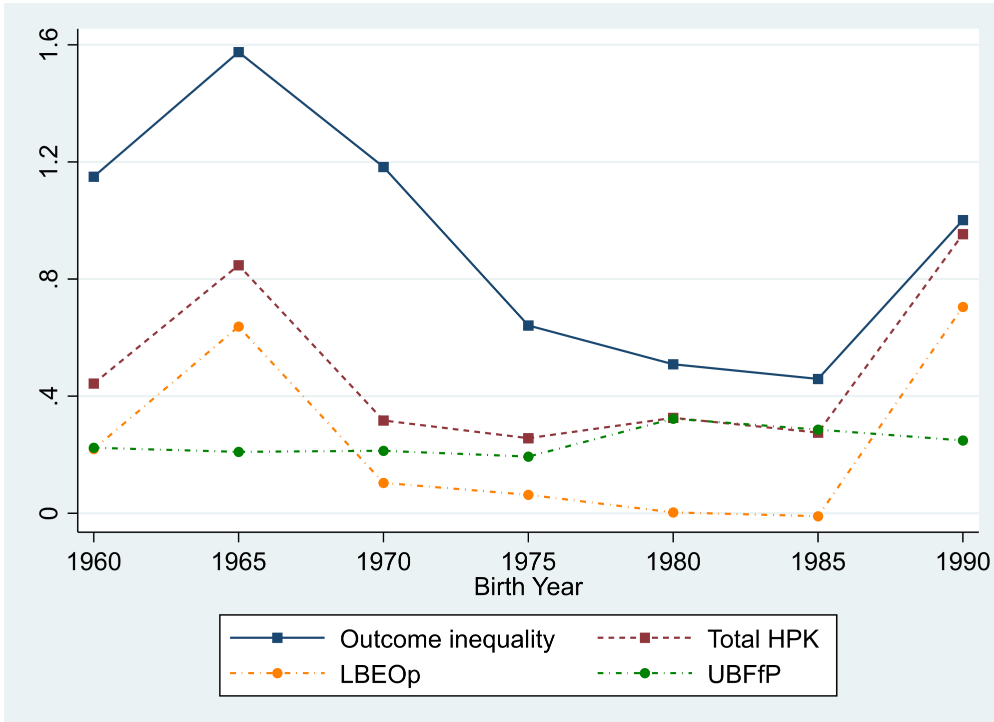

Figure 3.3.3: HKP Unfair inequality and decomposition: South Africa

In Fig 3.3.3, the green line shows UBFfP and the orange line shows LBEOp. We first notice a increasing trend of LBEOp, which is in line with our findings for the Romerian unfair inequality in Fig 1. UBFfP remains quite stable and show a slight decreasing trend over the years. UBFfP is smaller compare to LBEOp, which indicates that there is more inequality violating EOp than FfP.

Overall, there is a increasing trend of unfair inequality in South Africa for the younger cohorts, using 1975 as a separating point. The magnitude of unfair inequality is also fairly significant. Hufe, Kanbur, Peichl (2018) measured unfair inequality for US and European countries. The average absolute unfair inequality in 31 European countries is 0.029. As for US, it starts from a level of 0.023 in 1969, unfair inequality attained a level of 0.130 in 2012. Compare their results to the average unfair inequality in South Africa among the five waves, 0.33, its clear that South Africa has a larger extent of unfair inequality.

3.3.3 Comparison

Table 3.3.4 combines the unfair inequality measurement result for both the Roemerian method and the HKP method. We see that overall, HKP method yields a higher unfair inequality when we compare IOA with total HKP. According to the theoretical structure of these two methods, the exceeding part is unexpected since HKP include the principle of FfP, creating a longer distance for some poor to reach the poverty line instead of just the mean income of its type. In order to obtain enough observation for each type, we have to cut down the number of circumstance variables or make the existing ones more coarsely defined. This adjustment reduces the explanatory power of the circumstance. The parametric Roemer model doesnt have this problem of keeping a certain number of observations in a type cell. The fact that the by using the less finely defined circumstance, we still arrive at higher amount of unfari inequality indicates the importance of including the FfP principle. However, this difference doesnt change the trend of unfair inequality due to unequal opportunity that we observe. For both of the measurement, we observe a decreasing trend of unfair inequality stemming from unequal opportunity.

Although we can cross-validate the trend of EOp using the two methods, the features of unfair inequality violating FfP can only be obtained using HKP method. From Table 3.3.4 Panel C, we find that unfair inequality to FfP takes nearly 84.5% of the total HKP unfair inequality. This is not a small size at all. Bringing in the HKP method certainly generate a new dimension when viewing unfair inequality and suggests a shift of the recent focus of attention from the uppermost parts of the income distribution to the lower percentiles.

One thing in common for these two methods is that they still yield a lower bound of the unfair inequality of the society if we couldnt include all the proper circumstances that individuals face. Nevertheless, the Romerian method and the HKP method can be a good complement to each other when we want to get a fuller picture of the how much of the inequality is unfair and how it changes.

Table 3.3.4: NIDS: Comparison of Roemerian Method and HKP method

	All	58-62	63-67	68-72	73-77	78-82	83-87	88-92
Panel A: Outcome inequality								
GE(0)	1.49	1.15	1.58	1.18	0.64	0.51	0.46	1.00
Panel B: Roemerian unfair inequality								
IOA	0.24	0.51	0.31	0.17	0.25	0.16	0.1	0.62
IOR	15.88%	44.33%	19.70%	14.54%	38.32%	30.86%	21.58%	61.53%
Panel C: HKP unfair inequality								
HKP	0.33	0.44	0.85	0.32	0.26	0.33	0.28	0.95
Share of HKP	22.36%	38.53%	53.79%	26.80%	39.97%	64.02%	60.03%	95.22%
UBFfP	0.28	0.22	0.21	0.21	0.19	0.32	0.29	0.25
LBEOp	0.05	0.22	0.64	0.1	0.06	0	-0.01	0.7

3.4 Conclusion

This paper presents an up-to-date nationally representative analysis of South Africas unfair inequality in individual income using data from the National Income Dynamics Study (NIDS). Unfair inequality is measured in two ways. We first focus on the unfair inequality that violates the normative principle of Equal Opportunity (EOp). Following the conceptual contribution of Roemer (1998), we use an ex-ante parametric model to examine to what extent individuals' incomes inequality is due to circumstance factors, that is, factors beyond their control. This method allows us to calculate the contribution of each circumstance by comparing the outcome when we take each circumstance out of the model one by one. However this single definition of fairness is not enough. Another essential normative principle needed is satisfying the basic need. Therefore, we take a step further to measure the inequality that violates both EOp and Freedom from Poverty FfP based on Hufe, Piechl and Kanbur (2018). This method allows us to decompose the unfair inequality into its EOp and its FfP components.

The first measure of unfair inequality (we also call it Roemerian unfair inequality) indicates that in average 14.7% of the total outcome inequality is unfair. After we separate the samples into seven cohorts according to their birth year, we further observe an increasing trend of this Roemerian unfair inequality for the younger cohorts. Among all the circumstance variables we include, race and parental education significantly dominate all others by explaining 23% of the unfair inequality each. Our second measure of unfair inequality, which we called HKP unfair inequality, takes around 22.36% of the unfair inequality in our baseline result when the food poverty line is used. By further decomposing HKP unfair inequality into its EOp and its FfP components, we find 84.5% stems from violating FfP and 15.5% stems from violating FfP. Over the cohorts, HKP unfair inequality first decreases and then increase as the cohort members get younger.

Our results reflect a high degree of unfair inequality in South Africa. Inequality of opportunity is substantially correlated with racial inequality. Younger people are facing higher unfair inequality in the recent years, especially unequal opportunity. The impact of race and parental education in individual income has been increasing, with an decreasing influence from gender. Given that race and parental education are among the most influential of circumstances for the older cohorts, policy interventions aimed at reducing the racial pay gap and investing on young adult's education should be considered in order to create more equal opportunity. A large portion of inequality stemming from violating FfP reflects that basic social security networks should also be considered.

REFERENCES

Bourguignon, F., Ferreira, F. H., & Menéndez, M. (2007). Inequality of opportunity in brazil. *Review of income and Wealth*, *53*(4), 585–618.

Bourguignon, F., Ferreira, F. H., & Walton, M. (2007). Equity, efficiency and inequality traps: A research agenda. *The Journal of Economic Inequality*, *5*(2), 235–256.

Brunori, P., Ferreira, F. H., & Peragine, V. (2013). Inequality of opportunity, income inequality, and economic mobility: Some international comparisons. In *Getting development right* (pp. 85–115). Springer.

Checchi, D., Peragine, V., & Serlenga, L. (2010a). Fair and unfair income inequalities in europe.

Checchi, D., Peragine, V., & Serlenga, L. (2010b, June). *Fair and Unfair Income Inequalities in Europe* (IZA Discussion Papers). Institute for the Study of Labor (IZA).

Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, *99*(4), 906–944.

Dworkin, R. (1981). What is equality? part 2: Equality of resources. *Philosophy Public Affairs*, *10*(4), 283–345.

Ferreira, F. H., & Gignoux, J. (2008). *The measurement of inequality of opportunity: Theory and an application to latin america*. The World Bank.

Ferreira, F. H., & Peragine, V. (2015). *Equality of opportunity: Theory and evidence.* The World Bank.

Fleurbaey, M. (1995). Three solutions for the compensation problem. *Journal of Economic Theory*, *65*(2), 505–521.

Fleurbaey, M. (2008). *Fairness, responsibility, and welfare*. Oxford University Press.

Golley, J., & Kong, S. T. (2018). Inequality of opportunity in china's educational outcomes. *China Economic Review*, *51*, 116–128.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674.

Hufe, P., Kanbur, R., & Peichl, A. (2018). Measuring unfair inequality: Reconciling equality of opportunity and freedom from poverty.

Hufe, P., Peichl, A., & Daniel, W. (2019). Lower and upper bounds of inequality of opportunity in emerging economies.

Kanbur, R., & Wagstaff, A. (2014). *How useful is inequality of opportunity as a policy construct?* The World Bank.

Lefranc, A., Pistolesi, N., & Trannoy, A. (2008). Inequality of opportunities vs. inequality of outcomes: Are western societies all alike? *Review of income and wealth*, *54*(4), 513–546.

Leibbrandt, M., Woolard, I., McEwen, H., & Koep, C. (2010). Employment and inequality outcomes in south africa. *University of Cape Town: Southern Africa Labour and Development Research Unit*.

Niehues, J., & Peichl, A. (2014). Upper bounds of inequality of opportunity: theory and evidence for germany and the us. *Social Choice and Welfare*, *43*(1), 73–99.

Piraino, P. (2015). Intergenerational earnings mobility and equality of opportunity in south africa. *World Development*, *67*, 396–405.

Ravallion, M. (2017). A concave log-like transformation allowing non-positive values. *Economics Letters*, *161*, 130–132.

Rawls, J. (1958). Justice as fairness. *The philosophical review*, *67*(2), 164–194.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Roemer, J. (1998). *Equality of opportunity*. Cambridge: Harvard University Press.

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner.

Trannoy, A., Tubeuf, S., Jusot, F., & Devaux, M. (2010). Inequality of opportunities in health in france: a first pass. *Health economics*, *19*(8), 921–938.

Van De Gaer, D. (1995). Equality of opportunity and investment in human capital.

Zhang, Y., & Eriksson, T. (2010). Inequality of opportunity and income inequality in nine chinese provinces, 1989–2006. *China Economic Review*, *21*(4), 607–616.

APPENDIX A

REFERENCE INCOME CONSTRUCTION

Our baseline reference income distribution that follows both EOp and FfP is constructed as follows:

$$y^r_{i\,EOp+FfP} = \begin{cases} y_{min}, & \text{if } \mathrm{y}^e_i < y_{min} \\ y_i\Big[1 - \tilde{y}_i\Big(\tau^{FfP} + \tau^{EOp}_t(1 - \tau^{FfP})\Big)\Big], & \text{otherwise} \end{cases}$$

Where $\tilde{y}_i = \frac{y_i - y_{min}}{y_i}$, $\tau^{FfP} = \frac{N_P(y_{min} - \mu_P)}{N_R(\mu_R - y_{min})}$ and $\tau^{EOp}_t = \frac{\mu_t + \frac{N_{P\cap t}}{N_t}(y_{min} - \mu_{P\cap t}) - \tau^{FfP}\left(\frac{N_{R\cap t}}{N_t}(\mu_{R\cap t} - y_{min})\right) - \mu}{\mu_t + \frac{N_{P\cap t}}{N_t}(y_{min} - \mu_{P\cap t}) - \tau^{FfP}\left(\frac{N_{R\cap t}}{N_t}(\mu_{R\cap t} - y_{min})\right) - y_{min}}$

Now lets consider the case sf we only follow the principle of EOp. This will lead to $y_{min} = 0, N_P = 0, N_R = N$ and $N_{R\cap t} = N_t$. Hence $\tau^{FfP} = 0, \tau^{EOp} = \frac{\mu_t - \mu}{\mu_t}$ and our reference income distribution would look like:

$$y^r_{i\,EOp} = y_i \cdot \frac{\mu}{\mu_t}$$

On the other hand, if we only focus on FfP, we only lift the people above the poverty line instead of considering the distribution among the rich. In this case $\mu_t = \mu, N_R \cap t = N_R, \mu_R \cap t = \mu_R$ and hence $\tau^{EOp} = 0$. Consequently, the reference income would be constructed as follows:

$$y^r_{i\,FfP} = \begin{cases} y_{min}, & \text{if } \mathrm{y}^e_i < y_{min} \\ y_i\Big[1 - \tilde{y}_i \cdot \tau^{FfP}\Big], & \text{otherwise} \end{cases}$$

APPENDIX B

DESCRIPTIVE ANALYSIS FOR COHORTS

B.1 China

Table B.1.1: Descriptive Statistics over Cohort (CFPS)

Birth year	55-57	58-60	61-63	64-66	67-69	70-72	73-75	76-78	79-81	82-85
Avg age (Yrs)	53.98	51.08	47.75	45.02	42.02	39.07	36.13	33.01	30.05	27.09
Avg income (RMB/year)	7940	8827	10400	10878	11177	12073	13650	14235	15153	15623
Han (%)	93	93	94	93	92	92	90	88	89	91
Male (%)	49	51	45	47	0.46	47	46	46	49	48
Urban Hukou (%)	14	16	12	9	10	9	10	12	14	17
Born Inland (%)	49	46	54	56	57	59	61	59	55	53
Parent(s) ever educated (%)	36	41	43	47	54	62	67	71	78	80
Parent(s) with non-ag job (%)	69	65	67	72	69	69	71	63	63	58
Poverty rate (head count ratio)	0.42	0.33	0.33	0.33	0.31	0.3	0.3	0.29	0.29	0.26
Population	1518	1187	1554	1850	1798	1801	1447	1222	1133	1133

B.2 South Africa

Table B.2.1: Descriptive Statistics over Cohort (NIDS)

Birth year	58-62	63-67	68-72	73-77	78-82	83-87	88-92	Total
Age	52.28	48.12	43.45	38.58	33.62	29.35	26.35	38.87
wage	11691.3	60026.2	20110.9	11887.9	6684.6	6788.6	7364.6	18030.2
Female (%)	53.92	54.88	56.98	55.21	49.97	50.23	43.32	52.64
African (%)	75.48	75.23	74.62	74.57	80.88	84.46	85.22	78.32
Parents no schooling (%)	54.73	53.02	47.09	43.22	34.47	23.95	21.21	40.05
Parents w/ no or elem job (%)	68.46	63.39	63.33	58.30	58.69	52.43	51.41	59.46
Poverty rate (head count ratio)	0.56	0.60	0.62	0.63	0.69	0.70	0.73	0.65
Population	983	1292	1497	1585	1433	1499	778	9067

APPENDIX C

CONTRIBUTION OF EACH CIRCUMSTANCE OVER COHORTS

C.1 China

Table C.1.1: Contribution of different circumstances, cohort pattern (CFPS)

Cohort	Ethnicity	Hukou	Gender	Birth place	Parental Educ	Parental Occup
55-57	0.00%	4.52%	51.97%	13.21%	11.30%	2.03%
58-60	5.93%	3.50%	57.06%	12.42%	8.35%	1.96%
61-63	0.00%	14.00%	64.14%	7.45%	3.62%	1.48%
64-66	0.37%	13.42%	53.09%	3.06%	6.56%	2.81%
67-69	0.08%	5.75%	56.16%	4.32%	8.48%	3.78%
70-72	0.56%	11.07%	46.82%	7.31%	10.59%	1.72%
73-75	0.92%	7.48%	42.31%	12.18%	10.63%	1.80%
76-78	4.73%	1.28%	45.53%	5.66%	1.89%	10.49%
79-81	1.10%	3.68%	40.91%	9.73%	9.42%	8.25%
82-85	1.94%	9.25%	36.38%	14.65%	9.06%	0.16%

C.2 South Africa

Table C.2.1: Contribution of different circumstances, cohort pattern (NIDS)

Birth year	Gender	Race	Parental job	Parental educ	Cohort
All	9.44%	22.88%	1.40%	23.01%	4.01%
58-62	12.12%	25.57%	2.73%	10.17%	-
63-67	8.78%	10.29%	-1.57%	23.25%	-
68-72	25.12%	2.17%	7.76%	30.41%	-
73-77	11.59%	13.92%	0.74%	28.21%	-
78-82	3.55%	45.12%	3.71%	26.41%	-
83-87	2.24%	28.90%	1.68%	33.94%	-
88-92	6.59%	59.62%	3.53%	24.42%	-

APPENDIX D

INCLUDING 0 INCOME USING CONCAVE LOG-LIKE TRANSFORMATION

D.1 Kernel Density

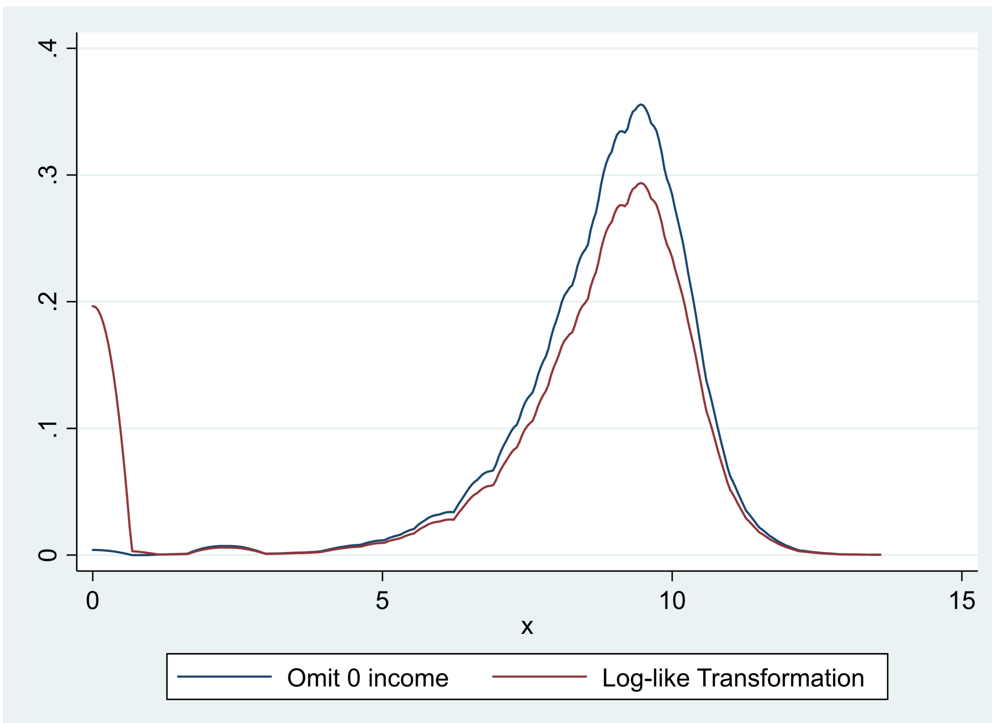

Figure D.1.1: C. Kernel density of before and after transformation

Note: kernel = epanechnikov, bandwidth = 0.2654
The blue line shows the distribution of the log form of the original income. (12,074 samples) The red line shows the distribution of after-transformation income that include 0. (14,622 samples)

D.2 Regression results after concave log-like transformation

Table D.2.1: Regression result after concave log-like transformation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	All	55-57	58-60	61-63	64-66	67-69	70-72	73-75	77-79	80-82	83-85
Han Chinese (YES = 1)	0.50***	-0.08	1.13**	-0.34	0.6	0.58	-0.16	0.99**	0.72*	0.44	1.05**
Hukou at age 3 (Urban = 1)	0.56***	-0.08	0.55	0.52	1.12***	0.51	0.59**	0.38	0.11	0.87**	1.04***
Male (YES = 1)	2.45***	3.01***	2.35***	2.30***	2.06***	2.24***	2.55***	2.44***	2.67***	2.68***	2.34***
Birth region: Baseline North											
East=1	0.68***	0.04	0.62	0.74*	0.92***	0.63*	0.68**	0.88**	0.67*	0.83**	0.70**
Mid=1	0.30***	-0.31	0.14	0.14	0.44	0.53	0.28	0.47	0.16	0.80**	0.22
West=1	0.63***	-0.03	0.93**	0.47	0.62*	0.59*	0.34	1.13***	0.39	1.13***	0.68*
Parental Education: Baseline Illiterate/Semi-literate											
Primary=1	0.34***	0.43	0.1	0.06	0.55***	0.41*	0.60***	0.72***	-0.25	0.26	0.32
Secondary=1	0.44***	0.80*	-0.07	0.19	0.47*	0.38	0.80***	0.67**	0.15	0.61**	0.03
College and above=1	1.42***	0.4	0.81	0.91	1.43***	1.45***	2.24***	1.11*	1.04*	1.53***	1.17*
Parental job: Baseline non-agricultural											
Agricultural=1	-0.49***	-0.24	-0.15	-0.78***	-0.26	-0.94***	-0.36*	-0.62**	-0.64**	-0.35	-0.36
Birth cohort	Yes	-	-	-	-	-	-	-	-	-	-
Constant	4.68***	5.33***	4.32***	6.40***	5.11***	5.79***	6.01***	4.70***	5.89***	4.90***	5.28***
Observations	14,622	1,509	1,175	1,571	1,839	1,796	1,798	1,449	1,221	1,135	1,138
R-squared	0.16	0.17	0.13	0.12	0.12	0.15	0.19	0.16	0.18	0.18	0.17

Note: All figures are weighted by NIDS sample weights to be nationally representative.
Further details on birth cohorts and robust standard errors available on request.
*** p¡0.01, ** p¡0.05, * p¡0.1